Practical guidance for evaluating and interpreting developmental toxicity tests

Stephen B. Harris[a,*], John M. DeSesso[b]

[a] *Stephen B. Harris Group, 6109 Madra Avenue, San Diego, CA 92120, USA*
[b] *Center for Environment, Resources and Space, The MITRE Corporation, McLean, VA 22102, USA*

Abstract

The complex physiological interactions among a pregnant mammal, her embryos, and their placentae pose considerable challenges to investigators who conduct safety tests for the assessment of potential developmental toxicity. Many individuals who review developmental toxicity safety tests are not trained in this specialized area of toxicology. This paper presents a concise introduction to the science that underlies developmental toxicology for those individuals. The purpose of the paper is to educate the reader about appropriate test procedures, the types of data that are collected, and evaluation of studies. To these ends, the paper explains important terminology and study designs; makes comments concerning what should be considered acceptable developmental toxicity data; and provides insights and rules of thumb regarding the evaluation and interpretation of the data.

1. Introduction

Mammalian embryonic development is a complex, yet wondrously orchestrated phenomenon. A knowledge of embryology is a prerequisite for understanding the mechanisms whereby developmental toxicants interfere with developing embryos. However, many individuals who are responsible for reviewing developmental toxicity safety test reports do not possess a strong background in either developmental biology, embryology, or teratology. While reviewers of developmental toxicity test reports are not required to understand the mechanisms that cause the induced, adverse developmental effects, a knowledge of the aforementioned areas would add to their comfort level.

Critically reviewing and interpreting developmental toxicity test reports are significant challenges for a number of other reasons. Not only has the science underlying the assessment of potential developmental toxicity been incompletely codified, but also

---

* Corresponding author. Tel.: (+ 1-619) 469-7886.

many important issues relative to the interpretation and extrapolation of animal data are unresolved. Even such apparently basic items as the definition of malformations, variations, and anomalies have not been agreed upon. A *nomina terata* is not available and, therefore, the names of structural malformations and variations vary dramatically from laboratory to laboratory. Consequently, it is the purpose of the present paper to briefly introduce the science underlying regulatory developmental toxicology to non-developmental toxicologists.

The primary concerns of developmental toxicology reports are the identification of substances that are potentially hazardous to developing organisms and the establishment of a developmental toxicity no observable adverse effect level (NOAEL) as the first step in the assessment of human developmental toxicity. While it is paramount that the reports of such studies be well documented, it has been the experience of the authors that many reports of developmental toxicology studies are of poor quality. This lack of quality is due to either incorrectly performed studies or inadequate reporting of the study. In both cases, additional time and money are required to either fix the report (if possible) or repeat the study.

Because many individuals who receive developmental toxicology reports for review are not trained in this specialized area of toxicology and, thus, are unprepared to critically review the study protocols and final reports, we have included brief descriptions of, and commentaries on, the underlying assumptions, basic experimental design, kinds of data that are collected, and rules of thumb for interpretation of conventional developmental toxicity studies. Our objective is to provide guidance for practical evaluation of developmental toxicology studies based on our experience in this area of testing. More complete discussions of the theory and requirements of developmental toxicity testing can be found in documents published by others, including several regulatory agencies [1–10].

In order to understand the rationale that underlies the experimental design and kinds of data that are collected in developmental toxicity studies, the uniqueness of the pregnant mammal as an experimental system must be understood. This specialized test system is composed of three interdependent, functional units: the pregnant dam, the placenta, and the embryo. In animals that produce multiple young, each embryo has its own placenta. To reach the embryo, a test substance administered to the pregnant dam must traverse the placenta. Thus, each of the functional units may be the target for a test substance and each of the functional units may be able to metabolize substances that pass through it. In general, administration of test substances in a developmental toxicity test does not begin until implantation is completed. Depending upon the species involved, five to eight days elapse between fertilization of ova in the upper female reproductive tract and the implantation of the embryo into the uterine wall with the concomitant development of the placenta (see Table 1). In the cases of the mouse and the rabbit, treatment is begun prior to the completion of implantation because the embryos of those species begin major organogenesis (the period of peak sensitivity to many developmental toxicants) before implantation is finished.

Another unique feature of developing mammals as test animals is that the embryo continually changes both morphologically and biochemically/metabolically during

Table 1
Comparative temporal landmarks and testing schedules for developing mammals (in gestational days[a])

Species	Gestational milestones			Typical developmental toxicity schedules	
	Implantation ends	Organogenesis ends	Parturition	Exposure period	Sacrifice of cesarean section
Hamster	4.5–5	13	16	5–14	16
Mouse	7	15	19–20	6–15	18
Rat	5–6	15	21–22	6–15	21
Rabbit	7.5	18	31–33	6–18 or 7–19	29
Guinea pig	6	~ 29	64–68	6–30	60
Monkey	9	~ 44–45	166	9–45	100
Human	6–7	~ 50–56	266	NA[b]	NA

[a] Day of confirmation of mating = gestational day 0.
[b] NA = not applicable.

gestation. During the period when the rudiments of the major organ systems are laid down (organogenesis), the embryo is maximally sensitive to agents that may cause birth defects and altered growth. Organogenesis begins near the time of implantation. Its total duration differs from species to species, but it is positively correlated with the duration of gestation for the particular species (see Table 1). In most developmental toxicity test designs, the exposure of pregnant females to test agents begins near the completion of implantation and continues through the period of organogenesis.

Developmental toxicity studies determine the potential of an agent administered to a pregnant mammal to induce adverse effects on her developing offspring. While data from the pregnant animal are collected throughout the study and analyzed in the final report, the four major endpoints of developmental toxicity studies relate to the offspring. These endpoints include the death of developing organisms, structural abnormalities in offspring (congenital malformations), altered growth, and functional deficits. All four manifestations are considered a concern; a biologically significant increase in any of them is considered indicative of an agent's potential to perturb development and produce a developmental hazard. The standard developmental toxicity tests examine the effects of test compounds on the first three manifestations. Functional deficits seldom have been evaluated in routine testing, although recently the developmental toxicity assessments of some substances have included functional evaluations [11–14].

The health status or physiological well-being of pregnant females may affect the offspring (see discussion in [15]). Prominent among "maternal factors" that may adversely affect the offspring is stress [16, 17]. In addition, agent-induced toxic effects in the pregnant female may elicit indirect effects in the offspring. Thus, while developmental toxicity is often described as a stand alone entity, it is intimately related to maternal toxicity. Indeed, it is often difficult to distinguish effects mediated through toxicity in the mother from those caused by direct action within the embryo itself [18, 19]. Consequently, it is imperative [1] to minimize extraneous maternal factors

that could affect the outcome of pregnancy; and [2] to utilize a range of doses including a high dose that elicits maternal toxicity.

We have made an effort to organize the remainder of this paper according to topic areas related to the review of developmental toxicity reports. We present comments that relate to each topic area with regard to what should be considered acceptable data, as well as insights for the evaluation and interpretation of the data.

2. Evaluation of developmental toxicity test reports

2.1. General considerations

Type of study

There are two types of developmental toxicity studies: the range-finding (pilot) study and the definitive developmental toxicity (segment II) study. The purpose of a range-finding study is to establish the dose levels for the definitive developmental toxicity study. Typically, range-finding studies employ more dose levels and fewer animals per group (we recommend 4–6 dose groups with 8–10 pregnant rodents or rabbits per group). Range-finding studies seek to determine a dose of test substance that elicits minimal maternal toxicity (to be used as the high dose in the definitive study) and a dose that causes no adverse effects in the offspring. The amount of in-life maternal data collected is similar to that in the definitive developmental toxicity study, but the data collected at cesarean section are usually limited to the gross examination and weighing of fetuses.

The primary purpose of the definitive developmental toxicity study is to determine whether the test agent induces any adverse effects on the developing organism and, if so, to establish the NOAEL. A great amount of post-mortem data is collected in the definitive developmental toxicity study.

Study protocol

The protocol describes in detail the plans for a study, including the test species, dosage levels, mode of exposure, number of animals per group, and the observations that are to be made. The study should have well-defined maternal and fetal observations. Furthermore, all methods of fetal examination should be clearly stated. If they are not adequately described in the protocol, they should be available in the standard operating procedures of the laboratory. It is imperative these procedures are spelled-out.

The protocol should conform to guidelines and testing requirements of the appropriate regulatory agencies, however, the testing requirements are minimum data needs. Additional testing or modification of routine study designs is sometimes necessary for the assessment of developmental toxicity potential. For instance, deviations from basic protocols are acceptable with proper reasoning, e.g., a postnatal phase may be necessary to distinguish dilated renal pelvis (which is a reversible condition) from true hydronephrosis (a kidney malformation).

Accurate records should be maintained and all experimental data should be quality-assured. This is accomplished by conducting data inspections and audits

according to the Good Laboratory Practices (GLPs) regulations mandated by the US Food and Drug Administration [20] and those subsequently developed by the European Chemical Industry Ecology and Toxicology Center [21], the Organization for Economic Cooperation and Development [22] and the US Environmental Protection Agency [23, 24]. The intention of these regulations is to ensure the quality and integrity of the data, but not to limit informed scientific judgment when the data may be incomplete. Compliance with GLPs facilitates reconstruction of a study and provides a framework for practicing good science.

Final reports

All reports submitted as final must be signed and dated by both the Study Director and the Director of Quality Assurance. If a report is not signed and dated, reviewers should assume that it is subject to change and does not represent the final position of the laboratory. Draft reports are not acceptable for fulfillment of regulatory requirements.

Presentation of maternal and fetal findings

Clear summary and individual table formats should be used to facilitate easy audit and review. It must be possible to associate all reported maternal and fetal findings with individual animals. The data must be presented in a manner that allows the identification of those females that showed any given clinical signs on any given day and to identify individual fetuses that presented with each variation and malformation. All reported mean data should be carefully compared to submitted individual data for possible inconsistencies. The appropriate application of statistical methods should be verified.

2.2. Animals

Appropriate test animals

It is important to obtain animals that possess a uniform genetic background, are disease-free, and of similar reproductive age and parity. Nulliparous (virgin) females are required for testing because confirmation of pregnancy in previously pregnant females cannot be accurately determined. Basic animal husbandry practices are required in the laboratory [25] and should conform to guidance published by the US Department of Agriculture [26, 27].

Choice of species

Testing is required in two species to support the registration of a product intended for food use (i.e., when tolerances or exemptions from tolerances are considered) and for nonfood uses, if significant exposure of women of child bearing age may reasonably be expected. The usual test species are one rodent (rat or mouse) and one nonrodent (rabbit). Studies conducted using species other than these may be acceptable, but justification of use is required. The "most appropriate" animal species (i.e., the species that reacts to the test substance most like humans) is used to estimate risk. In the absence of knowing which species is most appropriate, the most sensitive species

is used. This is based on the premise that for proven human developmental toxicants, humans are at least as sensitive as the most sensitive animal species [28].

2.3. Test compound and dosing

Test compound

Impurities in the test material may be an important factor in the teratogenic potential of a compound. In some cases, the impurities may be the sole cause of the adverse effects. Consequently, information regarding purity of the test compound and identification of impurities should be available in the final report or the compound registration. In most cases, the technical material intended for commercial use is tested and, consequently, testing of the formulation is not required.

Dosing formulations

The test compound is usually mixed with either vehicle, drinking water, or feed prior to administration to the test animals. It is important that the concentration of test material be accurate. Predosing and postdosing chemical analyses should be performed to confirm the concentration of the dosing formulations. One problem in developmental toxicology studies occurs when the administered dose of test substance is not the intended dose. This is especially true when the target concentration (nominal) of the dosing formulation presented in the protocol is not the same as the analyzed concentration (analytical). Dosing formulations should be within a range of $\pm 10\%$ of the target concentration. If the analytical dosing concentrations are outside this range, the study should be rejected.

Vehicle

If a vehicle is used to deliver the test substance, the vehicle without the test substance should be given to the control group. The vehicle should not cause maternal or developmental toxicity. However, if there are inadequate data regarding the potential toxicity of the vehicle, the rationale underlying the choice of vehicle should be provided. A sham and an untreated control group may be warranted if the toxic properties of the vehicle are not known.

Route of exposure

The route of exposure chosen for developmental toxicity studies should be the same as the likely route for human exposure.

2.4. Experimental design

Dose selection

Doses for the definitive developmental toxicity study are selected based on the results of the range-finding study. Unless limited by the biological, physical or chemical characteristics of the test substance, the highest dose level should produce some overt maternal toxicity, such as statistically significant reduction in maternal body weight or body weight gain, but not more than 10% maternal deaths. Dose

levels that produce excessive maternal toxicity may result in an unexpected number of maternal deaths and abortions. Thus, results such as these will limit the usefulness of the study, and the laboratory may have to repeat the study. A study may also have to be repeated if the highest dose investigated causes no significant maternal toxicity. Optimally, the high dose should elicit mild maternal toxicity and the low dose should cause no adverse effects on the offspring.

Treatment groups

The number of treated groups should be sufficient to establish a dose–response relationship. A minimum of three treated groups at different dose levels and a concurrent vehicle-treated control group should be used. An adequate number of animals per group is essential to provide statistical power for the results. The required number is normally 20 pregnant rodents (we recommend 30 pregnant mice or hamsters; 25 pregnant rats) and 12–16 pregnant rabbits (we recommend 20 pregnant rabbits). In order for body weight data to be useful as a potential indicator of toxicity, test animals should be randomized such that all dose groups start with similar mean maternal body weights and variance. Although not required by regulators, a concurrent positive control group may be warranted if the laboratory performing the study is inexperienced.

Exposure period

The most commonly used study designs [2, 4, 7, 28] include timed-mating of healthy laboratory animals. The usual reference for timing of gestation is to denote as gestational day 0 the day that either (1) a vaginal plug is observed (in rats or mice), or (2) sperm is discovered in the vaginal lavage (in rats), or (3) that mating was observed (rabbits), or (4) that artificial insemination was performed (rabbits). Dosing of presumed pregnant dams extends throughout the period of major organogenesis (days 6–15 for rats and mice; 6–18 or 7–19 for rabbits; and 5–14 for hamsters; see Table 1). In experiments that require animals to be dosed by technicians (e.g., gastric intubations), dosing should be performed at the same time each day with not more than two hours elapsing between dosing of the first and last animals, if possible. The timing of exposure is a very important consideration in developmental toxicology studies because embryologic events occur during very narrow windows of time during gestation. This is especially true in species with short gestations, such as those that are normally used in routine developmental toxicity studies.

Extended exposure regimens

Occasionally, the exposure of pregnant animals may begin at the end of implantation and continue throughout gestation. Such extended exposure regimens may disclose developmental changes that would not have been detected under the exposure conditions of standard developmental toxicity studies. For instance, continued treatment of dams from the end of organogenesis to cesarean section generally results in a higher incidence of growth retardation, which is often characterized by decreased mean fetal body weights. In addition, the heart, brain, lungs, and gonads are organs that continue functional and morphological development after the end of

organogenesis. If an extended dosing regimen is used, these organs may show structural alterations that would not have occurred under standard exposure conditions. Thus, the length of the exposure period may affect the findings of a study and should be carefully stated when the results of the study are being interpreted.

Concurrence of test groups

All experimental groups (both treated and control) should be run concurrently. Staggering of the induction of pregnancy within dose groups is acceptable; however, the mean time of induction and pregnancy must not differ significantly from one dose group to another. Prolonged periods before achieving the number of presumed pregnant females in the study may indicate a mating problem caused by such factors as poor health among the animals or a stressful environment in the animal facility.

Termination

Sacrifice of the females is scheduled just prior to delivery in order to prevent cannibalization of malformed young (see Table 1).

2.5. In-life procedures

Animals should be handled at all times with good animal husbandry practices [25–27]. Observations for mortality, moribundity, and clinical signs are usually conducted once daily at the time of weighing, or at additional times if the test material is known to be toxic.

2.6. In-life data

Maternal deaths and abortions

Maternal death and/or abortions may be caused by factors other than the test substance. These additional possible causes of maternal death and/or abortion include diseases, environmental factors, and technical errors (e.g., mishandling of the animals). In the case of maternal death, the necropsy records should be examined to determine a plausible explanation for the death. For instance, the presence of inflammation (reddening) of the tracheal lining, congested lungs, nasal discharge, and the accumulation of fluid in the lungs are suggestive of either a technical error (i.e., accidental intratracheal intubation) or disease. Another frequent occurrence in rabbits is the presence of hairballs in the stomach. Clinical signs that are observed in the presence of hairballs include alopecia (hair loss), anorexia (loss of appetite), diarrhea, and death. Therefore, caution must be used when interpreting maternal death as an endpoint of test compound-induced maternal toxicity because death may be a spontaneously occurring event that is unrelated to the toxicity of the compound.

Similarly, abortions and total litter resorptions may be induced by factors unrelated to the toxic effects of the test compound [16, 17]. Environmental stress due to such factors as excessive noise in the animal quarters, variations in light-dark cycles, and rough handling by technicians may cause abortions, especially in rabbits. Total litter

resorptions do occur in rabbits, but are more frequent in rodents (e.g., mice, rats, hamsters) which do not tend to abort.

Maternal body weights and body weight gains

Maternal body weights should be measured daily. Alternatively, dams may be weighed on the day of mating; on gestational day 5; daily throughout the exposure period; at 3–5 day intervals during the postdosing period; and at sacrifice. The maternal body weight gain for discrete segments of gestation (e.g., during treatment or throughout organogenesis) is usually a more sensitive indicator of maternal effects than either the final body weight at term or the total body weight gain over the entire period of gestation. The reason for the increased sensitivity of incremental body weight changes is that they are easily detected and are not masked by the "rebound" weight gain that often occurs in treated animals during the postdosing period.

Clinical signs

Clinical observations are important qualitative indicators of toxicity. A knowledge of the expected clinical signs of toxicity characteristic of the test compound should have been gained from the range-finding and other toxicological studies. Clinical sign data are objective observations (e.g., presence of tremors, excessive salivation, hunched posture in mice) that should be noted by well-trained technicians. Such signs are probably among the most reliable criteria of maternal toxicity. In some instances, clinical sign data may provide a more sensitive indicator of maternal toxicity than changes in maternal body weight.

Clinical sign data should include the identity of the observed effect as well as the time of onset, intensity, and duration of the effect. Clinical observations may include, but are not limited to, the presence of diarrhea, excessive salivation and mastication, nasal or ocular discharges, loss of hair (alopecia), tremors, convulsions, coma and death. Additionally, alterations in the rate of respiration, alertness, posture, movement within the cage, consumption of food and water (see below) color of mucous membranes, color of urine, and frequency of urination should be recorded. Reported changes in behavior noted during daily observations of animals (e.g., animals that appear aggressive or depressed) are not as objective as clinical sign data. They are similar to symptoms reported by human patients. Both observed changes in behavior and symptoms require a subjective interpretation by the reporter. While changes noted in daily behavior of the animals should be noted, at present it is not possible to determine whether changes in behavior are early manifestations of subclinical toxicity that may be manifested as external clinical signs at higher doses.

Food and water consumption

Food and water consumption should be measured daily in those studies in which the test material is administered in feed or drinking water (in order to calculate the dose), or when appetite or excretory effects are suspected. Alterations in food and water consumption after initiation of dosing are endpoints that can be used to determine maternal effects. It must be borne in mind, however, that when the test

substance is given in either the diet or drinking water, consumption may be reduced due to unpalatability.

2.7. Necropsy procedures – general

Procedures for examining offspring have been published by several authors [8, 10]. After the females are sacrificed, the uterus is removed with the ovaries intact and weighed. Corpora lutea on each ovary are recorded for both rodents and rabbits. While it is difficult to distinguish luteal tissue from ovary in mice, pending EPA guidelines [4b] require the counting and recording of corpora lutea in this species. The uterus is opened along the antimesometrial border and the contents are examined. The numbers and locations of implantation sites, resorptions, dead and live fetuses are recorded. Live fetuses are detached from the uterus, weighed, and evaluated for sex and any external malformations. This is the extent of offspring analysis required in a range-finding study. It is also an acceptable practice in a range-finding study to just record the numbers of implantation sites, resorptions and live and dead fetuses. In a definitive developmental toxicology study, half of the living fetuses are prepared for visualization of skeletal structures by staining with dyes specific for bone (alizarin red S) [29] or bone and cartilage (alizarin red S – alcian blue) [30]; the remaining fetuses are subjected to visceral examination by either dissection [8, 31, 32] or the free hand razor blade sectioning technique [33]. Alternatively, all fetuses may be subjected to fresh, visceral dissections, after which either they are all prepared for skeletal visualization; or half of the fetuses may be decapitated so that half of the heads may be prepared for free hand razor blade sectioning while the remaining heads and all bodies are stained for skeletal visualization.

2.8. Necropsy data – maternal

Confirmation of pregnancy

At the time of laparotomy, the uterus is examined for the presence of offspring and/or resorption sites. The pregnancy (conception) index is calculated by dividing the number of confirmed pregnancies in a particular group by the number of mated females. The pregnancy index is generally used to assess reproductive performance. Depression of this index may be an important indicator of a reproductive toxic effect if treatment begins prior to mating and implantation. Since treatment should begin after implantation is completed in developmental toxicity studies, a low pregnancy index may suggest maternal health problems, poor animal husbandry, or that dosing was mistakenly initiated prior to the completion of implantation.

Pregnancy indices are generally lower when pregnant animals, especially mice, are shipped from a supplier compared to pregnancy indices of animals bred in-house. Developmental toxicity studies will usually have to be repeated if there are statistically significant differences in pregnancy indices among groups.

Number of corpora lutea

Corpora lutea are the remnants of sites on the ovary from which ova were discharged during ovulation. Each corpus luteum discharged a single ovum. Thus, the

number of corpora lutea is the same as the number of ova that were available for fertilization. The difference between the number of corpora lutea and the total number of implantations (i.e., number of fetuses plus resorption sites) is termed pre-implantation loss. Increased pre-implantation loss should not be considered a compound-related effect in a standard developmental toxicity study because the effects occurred prior to the initiation of dosing. Reviewers should examine the data sheets to insure that dosing did not begin prior to the completion of implantation and that there is no evidence of environmental stress in the animal facility.

Gravid uterine weight/corrected maternal body weight

The intact, gravid uterus and ovaries are removed from the animal by first transecting the vagina at a point just inferior to the juncture of the uterine horns and then cutting the mesentery that connects these structures to the posterior body wall. The weight of the pregnant uterus plus ovaries is considered to be the weight of the "products of conception". This can be a useful parameter in cases where a particular animal that has few fetuses per litter is compared to another animal with many fetuses per litter. For instance, in control animals that have many more pups than expected in a litter, it is not uncommon to have an average fetal weight that is less than normal; conversely, in litters that have only a few (e.g., 2 or 3) fetuses, the average fetal weight often greatly exceeds the normal average fetal weight. Despite the fact that the mean fetal weights in the two extreme cases presented above may differ significantly from each other, it is often the case that the gravid uterine weights will not. Thus, the gravid uterine weight can serve as an indicator of how much body weight gain resulted from the pregnancy irrespective of the number of fetuses.

The corrected maternal body weight is the final body weight of the animal prior to sacrifice minus the gravid uterine weight. This parameter allows the reviewer to determine whether compound-induced adverse body weight changes were due to primary effects in the mother or the fetal-placental unit (products of conception).

Numbers of implantations. resorptions. living and dead fetuses

Death of offspring presents as resorptions and dead fetuses (fetal wastage), especially in rats and mice. Rabbits may either resorb or abort. Together, these endpoints are termed postimplantation loss and they are expressed as a percentage of the total number of implantations per litter. Embryonic or fetal death may be caused by direct lethal effects of the test substance, by lethal malformations (whether spontaneous or compound-induced), by maternal toxicity (whether compound-induced or due to disease), or by environmental stress. Determination of a dose–response relationship strengthens conclusions regarding the developmental toxicity of a test compound. Reviewers should attempt to recognize a preponderance of one type of postimplantation loss; such a finding may suggest a specific stage in development during which the test compound was toxic to the developing organism.

The number of viable fetuses per litter is also recorded. This endpoint, when compared to the number of implantations per litter, provides useful information because it is a measure of developmental toxicity that includes lethality to offspring during all stages of development.

Organ weights and clinical chemistry

Organ weights and clinical chemistry data are not required by current regulations. However, dose-related effects on absolute and relative (absolute organ weight divided by the corrected maternal body weight) organ weights can be useful in the assessment of maternal toxicity. For example, the liver usually shows early signs of toxicity such as induction of enzymes, fatty change, or hydropic change. These changes are generally associated with increased liver weights. Consequently, if the maternal liver weights are reported, they should receive careful consideration.

Clinical chemistry data, such as hematology and enzyme markers, are occasionally determined in developmental toxicology studies. As in the case of liver weights, when clinical chemistry data are reported and notable effects occur, such data may be useful for determining a maternally toxic effect, even in the absence of other signs, such as a decrease in food consumption and an increase in other clinical manifestations.

2.9. Necropsy data – fetal

Fetal weights

Fetal body weight is an important and sensitive endpoint in the determination of developmental toxicity. Decreased mean fetal body weights compared to control values are an indication of growth retardation (one of the major endpoints of developmental toxicity). Since growth retardation may affect only certain fetuses within a litter, the parameter that is assessed is the mean fetal body weight. A decrease in mean fetal body weight for a litter, in a litter that is comparable in number of fetuses to control litters, generally indicates an embryo/fetotoxic effect. This reduction in mean fetal body weight may be the only indicator of developmental toxicity. It must be remembered, however, that among animals which deliver multiple young, individual fetal body weights tend to be heavier in smaller litters (see *Gravid uterine weight* above). Furthermore, the mean fetal body weight of males is greater than that of females.

The long-term interpretation of reduced mean fetal body weights discovered in a developmental toxicity study is not clear cut. Modest weight reductions may be transient. In other words, postnatal increases in fetal weight, size, and maturation can eliminate any appreciable differences between treated and control pups. In other cases, the fetal weight reduction is permanent (i.e., offspring fail to recover after birth). The potential reversibility of growth retardation must be assessed by considering offspring growth and viability data from not only the developmental toxicity studies but also multigeneration reproduction studies (wherein the pups are allowed to mature), because little is known about the long term effects of fetal body weight reduction.

Extremely small fetuses are termed either "runts" or "stunted", and are classified as malformed young. The criteria used in designating growth retarded fetuses vary among laboratories. Typically, however, offspring whose body weights are two or three standard deviations below the mean control fetal body weight or 25–30% less than the historical mean control body weight are classified as runts.

Fetal examinations – general comments

Live offspring are usually examined for external, soft-tissue and skeletal malformations. Changes from the normal anatomical pattern of fetuses should be recorded. The changes are graded according to their severity by such terms as malformations, anomalies, or variations [34]. Generally, malformations are regarded as those anatomical changes that are so severe that they interfere with the life or well-being of the fetus (e.g., spina bifida, cleft palate, phocomelia); anomalies are slight anatomical changes that exhibit only a small degree of detriment to the fetus (e.g., fused or wavy ribs, absence of nails on paws); and variations are structural alterations that regularly occur in normal animals (e.g., bifurcated gall bladder in rabbits; asymmetric sternebrae in rabbits and rodents). The determination of whether a particular alteration is classified as a malformation, an anomaly, or a variation frequently depends upon the training, experience and competence of the observer. Inter-observer bias has been the cause of significant inconsistencies in the conclusions reached by different laboratories. The following sections offer some precautions regarding the classification of findings during fetal examinations.

Gross structural changes

Fetuses should be removed from the uterus and examined promptly. While external alterations in structure (e.g., missing digits, cleft lip, umbilical hernia) are readily discerned, it is essential that the examinations proceed expeditiously to avoid potential artifactual findings. For instance, mouse, rat, and rabbit fetuses that are left too long in the uterus may present with flexed wrists and ankles that are frequently mistaken for anthrogryposis (club paws). Similarly, fetuses that have been allowed to sit on the examining table for extended periods before post-mortem external examinations can develop hyperextended or stiff joints.

Care should be exercised in the handling of the gravid uterus and the fetuses. Rough handling of these prior to the external examination can cause technician-induced subcutaneous hemorrhages. Reviewers should be aware of these possible artifactual findings and should be especially cautious when these are the only signs of developmental toxicity or when they are reported in both the experimental and control groups.

External fetal examination should include recording of the sex of each fetus, in addition to any structural alterations. Although agent-induced effects on sex ratio (number of females: number of males) are quite rare, reviewers should be apprised that some compounds may preferentially affect a particular sex.

Skeletal changes

Many alterations in skeletal structure are so common that they are regarded as alternative normal patterns. Examples of alternative normal patterns include the presence of either 12 or 13 pairs of ribs in rabbits, 14 pairs of ribs in rodents, and reduced ossification of the fifth sternebra in rodents and rabbits. Other minor changes in skeletal patterns (variations) appear to result from transitory developmental delays. These variations present as findings such as incomplete ossification of sternebrae and phalanges, supernumerary ribs (permanent) and wavy ribs (reparable during

postnatal development). These variations do not appear to have an adverse effect on the affected fetuses. While the developmental significance of these variations has not been defined, supernumerary ribs in mice and wavy ribs in rats have often been related to nonspecific maternal toxicity. Although such skeletal variations are not regarded as harmful developmentally toxic effects, they may be indicative of maternal toxicity or stress and/or fetal toxic effects if a significant dose-related increase of a particular variant is observed above concurrent controls. In such cases, reviewers should consider consulting the laboratory's historical control data to ensure that the findings are outside the range of a larger population of controls.

Reviewers should note that harvesting of fetuses 12–24 h earlier than recommended (e.g., on gestational day 19 or 20 rather than on day 21 for rats) can also result in observations of reduced or absent ossification in these same skeletal elements that are frequently recorded as variations. Had such fetuses been harvested at the recommended time, it is likely that reduced ossification would not have been observed. The reason for this is that rodent fetuses complete ossification rapidly during the last 48 h of gestation. Thus, reviewers should be alert for possible spurious increases in developmental variations due to scheduling mistakes for sacrifice times.

Visceral (soft tissue) changes

Malformations can occur in the organs of the body as well as in the external form and the skeleton. Frequently observed visceral malformations include malformations of the heart and great vessels (e.g., ventricular septal defects, tetralogy of Fallot, transposition of the great vessels), brain (e.g., hydrocephalus), kidneys (e.g., agenesis of kidneys, polycystic kidney), diaphragm (e.g., diaphragmatic hernia), and other organs. Perhaps the most important qualification for evaluating visceral alterations is a background in the normal anatomy of the test species. For instance, it is important to realize that rabbits and mice have gall bladders, but rats do not. It is helpful to understand the normal shapes of the organs, their relationships to each other, and the range of normal patterns. Thus, reviewers should understand that the diaphragm is composed of both membranous and muscular portions, and the membranous portion can be more or less transparent. Some fetuses have been diagnosed as exhibiting diaphragmatic hernias only because the technician could see through it; the structure was not probed for the presence of a membrane. Soft tissue changes that have not been seen before in some species (e.g., rat) are not necessarily malformations when seen in others. For instance, rabbits frequently exhibit a ventral pancreas, accessory spleen, small gall bladder, or retroesophageal subclavian artery.

As in the case of all fetal examinations, the dissections should be performed gently and with care. Reviewers should be aware that the edges of organs are smooth. Malformed organs do not present with jagged edges. The presence of jagged edges on an organ is an indication that the organ was damaged by the technician during the evaluation. As in the case of the external examination, rough handling of fetuses can cause petechial hemorrhages on internal organs. Improper handling during the cutting of the umbilical cord can cause backflow of blood into the fetus resulting in either intraabdominal hemorrhage or what appears to be a hemorrhagic liver.

Although it is not absolutely necessary, a working knowledge of the embryology of the organs being examined is helpful. Changes such as bifurcated or duplicated gall bladders in rabbits are not malformations when attached to a single bile duct, because this can result from slight changes in the branching pattern of the hepatic diverticulum during development. Similarly, abnormal lobulations of the liver or accessory renal arteries are not visceral malformations because they also arise from slight changes in normal embryonic development. When in doubt about such findings, reviewers should not hesitate to contact individuals with experience in the area of developmental toxicology.

2.10. Interpretation of results

Maternal toxicity

Since the highest dose of the definitive developmental toxicity study should cause some maternal toxicity, it is important that the data collected during the in-life phase of the study be adequate to assess signs of maternal toxicity. The endpoints that are useful in determining the presence of maternal toxicity include maternal death and abortion/resorption, reduced maternal body weights and body weight gains, and the presence of clinical signs.

High doses of a test compound will often cause maternal deaths accompanied by an increased incidence of abortions/resorptions among surviving females, especially in rabbits. Although these are indications of maternal toxicity, dose levels that cause a large number of maternal deaths and/or abortions/resorptions are usually not preferred as the high dose in the definitive developmental toxicity study because there will be an insufficient number of offspring to examine for developmental effects. Optimally, mild maternal toxicity will be observed only at the highest dose in a definitive developmental toxicity study. With some compounds, however, the margin between the dose that causes maternal death or abortion/resorption and the dose that induces other, less severe forms of maternal toxicity (e.g., decreased body weight gain, tremors) may be small. In such cases, it is not uncommon to see maternal deaths or abortions/resorptions even at the lowest dosage that induces the mildest forms of maternal toxicity.

Maternal body weight and body weight gain data are sensitive indicators of toxicity that are often used as a basis for determination of the NOAEL for maternal toxicity in most species. The pregnant rabbit is an exception, however, because pregnant rabbits may lose weight during a normal pregnancy. Ideally, body weight gain (or percentage change in body weight) of all groups of animals should be similar during the predosing period. Any decrease in maternal body weight seen among treated groups during the treatment period may be due to either toxicity of the test compound or decreased food consumption.

Test substance-induced reductions in maternal body weights or body weight gains are generally dose related; however, there may be instances in which the low dose group is affected while the high dose group is not. This is especially true when there is excessive maternal mortality in the high dose group, eliminating sensitive animals and consequently reducing the number of animals for comparison. In such a case as this,

the lack of a dose–response does *not* imply the absence of a compound-related effect. Similarly, in groups that have experienced numerous resorptions per litter, maternal body weights are not a clear indicator of the presence or absence of a treatment-related effect on the maternal animal. Comparison of the starting body weight to the corrected maternal body weight (terminal body weight minus gravid uterine weight) removes the uterine weight variability and allows assessment of a possible compound-induced effect on the female alone.

Reduction in food consumption may be an indicator of maternal toxicity in rodents and rabbits. This may occur soon after the initial dosing or may require repeated dosing before becoming evident. An increase in maternal body weight (adaptation or "rebound" effect) during the post-dosing period is common after a depression of food consumption during the exposure period. When reviewing a report, it may be possible to determine whether the reduced food consumption is due to a maternally toxic effect of the test substance or to unpalatability of the food by calculating the food efficiency index (FEI) for each group. The FEI, calculated as the grams of food consumed per gram of body weight gained, is a measure of how effectively food is used by the animal (e.g., body weight gain). If the FEI is similar between treated and control groups, then a maternally toxic effect is unlikely and unpalatability is the probable cause for reduced food consumption. Alternatively, if time and money permit, another experiment could be performed in which feed intake would be measured in groups of animals presented with either control or treated diet.

Rabbits exhibit several characteristics that confound the determination of maternal toxicity. For instance, change in maternal body weight is difficult to assess in rabbits because of their inherent erratic body weight gains and losses during gestation. It is also not unusual to observe a reduction in food consumption during the last week of gestation because rabbits attend to preparation of their nest for kindling rather than eating. In addition to the reduction in food intake, rabbits often exhibit hair loss (alopecia) in the abdominal region during the same period (the hair is being used to construct a nest). Consequently, reviewers (1) must be aware that these changes occur normally and (2) should use both concurrent control and historical control data when evaluating maternal toxicity in rabbits.

Developmental toxicity

If the administration of a test substance is associated with a demonstrable increased incidence of any developmental toxicity endpoint compared to the spontaneous incidence (as determined primarily from concurrent, but also historical control data), the agent can be suspected of being a developmental toxicant. If the endpoint of concern is that of congenital malformations, the agent is a suspected teratogen. Since developmental toxicity is an important noncancer endpoint for risk assessment purposes, the determination of a causal relationship between the administration of the test substance and the production of the endpoint in question is crucial.

A major, potential confounding influence on the interpretation of developmental toxicity safety tests is the tendency for fetuses in the same litter to exhibit similar endpoints. This tendency is termed the "litter effect" and has been ascribed to the fact that all fetuses in a particular litter experience the same maternal environment as their

littermates, but a different maternal environment from the fetuses of other litters in the same treatment group. Another potentially difficult problem facing reviewers is that the large number of offspring that are typically evaluated in a developmental toxicity test may give a false conclusion concerning the statistical significance of the data if they are analyzed with the fetus as the sampling unit. Reviewers must determine whether or not a particular developmentally toxic effect is real. The following paragraphs will discuss several tools and approaches to this problem.

A variety of statistical methods has been used for evaluating developmental toxicity data [35–37]. The methods differ with respect to the choice of sampling unit, either the number of females treated or the number of fetuses. For developmental toxicity safety tests, the appropriate sampling unit is the number of treated females [1]. Different approaches have been designed to account for the varying numbers of fetuses per litter and for the "litter effect". Consequently, some statistical analyses evaluate fetal endpoints that are expressed as incidence per litter while others analyze the number of litters with a fetus (or fetuses) that exhibit a particular endpoint. The choice of statistical methods should be clearly stated in both the protocol and the final report.

Reviewers should be aware that statistical analyses alone will not furnish an appropriate determination of whether or not an agent should be considered a developmental toxicant. When statistical significance is accepted at a probability of $p \leqslant 0.05$, it is expected that 1/20 comparisons will exhibit statistical significance due to chance alone. Since a great number of observations are made and analyzed in developmental toxicity studies (e.g., all of the individual skeletal elements that are checked and all of the viscera that are examined), reviewers should not be surprised to discover one to several observations per study that attain statistical significance. Other factors must be considered to determine whether the statistically significant finding is cause for concern.

A major factor to be considered when determining developmental toxicity is whether the finding in question exhibits a dose–response. A positive dose–response in the presence of statistically significant results in the group(s) receiving higher dose(s) is strong evidence for developmental toxicity. Reviewers should also recognize that dose-related trends in the incidence of fetal effects may occur without attaining statistical significance when compared to control values. For instance, congenital malformations, such as cleft palate, may occur at a low, but dose-related, incidence in treated groups with none of the groups being statistically different from control values. In cases such as this, it is important to be aware of the rarity of the observed endpoint in the test species. If the endpoint is rarely observed, the dose-related trend is more important for determining developmental toxicity than if the endpoint occurs regularly among control animals (see *Use of historical control data* below).

A second factor to be borne in mind when interpreting developmental toxicity data is that the incidence of the major endpoints may be related to each other such that the presence of one precludes the presence of others. For instance, when embryolethal doses are reached, embryolethality increases at the expense of the other endpoints such as growth retardation and malformations. This can help to explain why an increase in malformations may exist in the low and/or mid dose groups, but not in the

high dose group if the high dose group experienced a large increase in post-implanta-tion loss. It should also be noted that in other cases, where no adverse fetal effects are observed at the low and mid dose and the high dose causes extensive post-implanta-tion loss, a potential teratogenic effect may have been masked. A lowering of the high dose might have resulted in malformed fetuses. The previous comment notwithstand-ing, reviewers should note that the mechanisms underlying resorptions and post-implantation are not always the same as those leading to malformations (see discussion in [8]). In cases of difficult interpretation, reviewers should not hesitate to contact individuals with experience in the interpretation of developmental toxicity studies.

A third factor to be considered in evaluating the potential developmental toxicity of a substance is whether any observed fetal endpoints occurred in the presence or absence of maternal toxicity. Direct effects of the test substance on the embryo are considered to be those observed in the absence of maternal toxicity. If fetal effects are observed in litters from females that exhibited significant maternal toxicity, the effects may have been caused indirectly. As mentioned previously, a low incidence of "nonspecific" variations and malformations (e.g., wavy ribs, retarded ossification of sternebrae and phalanges, reduced body weight) occur at maternally toxic doses. However, reviewers should recognize that adverse fetal effects may be produced at doses that are only minimally maternally toxic. Such findings should not be con-sidered secondary to maternal toxicity. Rather, they indicate that both the embryo and the mother are sensitive to the same dose of test agent. It should also be noted that the presence of maternal toxicity does not guarantee that adverse fetal effects will be observed; some substances cause pronounced maternal toxicity but exert no apparent effects on offspring.

The final factor for consideration when interpreting developmental toxicity data is that not all malformations are caused by test agents. Virtually any type of malforma-tion can arise spontaneously in any animal [34]. Furthermore, a given malformation can be caused by more than a single agent or condition. A background incidence of spontaneously occurring malformations exists for each test species. Reviewers should recognize that when malformations occur in the absence of a dose–response, they may be spontaneous in origin. It is helpful to have access to the laboratory's historical control data when rare malformations arise (see discussion below) to determine whether or not a given malformation has been encountered previously in the laboratory.

Use of historical control data

Accumulation and maintenance of historical control data are highly recommended. Due to the large number of endpoints collected in developmental toxicity studies, statistically significant differences between data from treated and control groups may occur by chance alone. This is especially true in cases where the control incidence of a particular endpoint is unusually low and the incidence of the treated high dose group is slightly greater than expected. In such a case, it is helpful to know what the range of incidence for the endpoint in question has been among control litters. If the incidence of the endpoint in the treated groups is within the historical range, the finding is probably due to chance. Thus, a laboratory's historical control data may

prevent a false assumption of biological significance based on chance statistical significance. Furthermore, historical control data can substantiate that developmental alterations occur spontaneously in either untreated or vehicle-treated control animals. Moreover, historical control data provide the investigator information concerning changes that may be due to genetic drift, changes in animals' diet, seasonal changes, and differences among technicians in the manner in which observations are made and recorded.

Predictive value of positive findings

The relationship among the four major endpoints of developmental toxicity is often nonlinear and may vary with increasing dose such that, at higher doses, death of the offspring may preclude the expression of the other manifestations. Furthermore, the production of a particular type of developmentally adverse effect in an animal species does not predict the same developmentally adverse effect in another species (including humans). Consequently, a biologically significant increase in any of the major endpoints is a concern.

It has been acknowledged that the predictive value of animal teratogenicity tests for inferring risk to humans is indeterminate [38]. Virtually every chemical known to be teratogenic in humans is teratogenic in at least one laboratory species; however, approximately 71% of substances for which there is human exposure information and which have been correlated with human adverse developmental effects are positive in a single species of test animal [39]. While the amount of human information is generally limited, and the shape of the dose–response curve for human developmental toxicity is not known, the previously mentioned lack of concordance between animal and human studies means that the predictability of a single animal study is unknown. Thus, while an adequately designed animal teratology study that provides a positive teratogenic response suggests the possibility of risk to humans, it is incapable of predicting whether that substance will, indeed, cause developmental toxicity in humans. Nevertheless, concern about possible human developmental toxicity is heightened when a given test agent causes developmentally adverse effects in more than one test species.

3. Adequacy of studies for regulatory purposes

An ubiquitous challenge of regulatory toxicology is the determination of safe levels of human exposure to toxicants. It is important to identify both those doses of test compound that produced adverse effects and those doses that did not. These effect levels are the lowest observed adverse effect level (LOAEL; the lowest dose at which there is a statistically and biologically significant increase in the frequency of an adverse effect compared to controls) and the NOAEL. For developmental toxicity risk assessment purposes, these two types of effect levels must be identified in both the pregnant animal and the offspring.

Minimal maternal toxicity should be present at the high dose level used in the definitive developmental toxicity study and a LOAEL and/or a NOAEL for maternal

effects should be determined. However, if a NOAEL for maternal toxicity cannot be determined, even at the lowest dose level, a repeat study is not necessary as long as a NOAEL for developmental toxicity (i.e., effects in offspring) can be determined from the study.

For offspring effects, any of the four manifestations of developmental toxicity can be used. As mentioned previously, standard developmental toxicity tests are designed to investigate death, malformations, and fetal growth retardation. Thus, the data should reveal any dose-related increases in post-implantation loss (i.e., resorptions), fetal body weight changes, and alterations in fetal anatomy. A NOAEL or LOAEL should be determined. The endpoint to be used is the most sensitive of the developmental effects. If a NOAEL cannot be determined from the study, and if the test animal used was the most sensitive species, a repeat study testing lower close levels will usually be required.

As previously mentioned, a threshold for developmental effects is assumed. It should be noted, however, that determining a NOAEL does not identify the threshold dose for developmental toxicity. Rather, it merely establishes that, under the conditions of the study, no adverse effects were observed at a particular dose level. The threshold is higher than the NOAEL but lower than the LOAEL. In many studies, the NOAEL may be an order of magnitude smaller than the LOAEL and, therefore, may be a poor estimate of the true threshold dose. Due to the limitations for risk assessment purposes of using a point estimate, like the NOAEL, as a surrogate for the developmental toxicity threshold dose, the EPA [1] is currently assessing other methods for quantifying dose–response relationships, such as the benchmark dose method [40, 41].

4. Conclusion

Developmental toxicity safety tests are a first step in the risk assessment of a chemical's potential developmental toxicity to humans. These tests are part of a specialized subarea of regulatory toxicology. Due to the complexities inherent to the test system (the materno-placental-embryo unit) and the unique experimental design, numerous parameters are evaluated as part of the determination of adverse effects in both the pregnant female and her offspring. The commentaries and insights provided above are meant to help educate non-developmental toxicologists about this very important discipline. The paper is a brief introduction to the discipline as it is encountered in developmental toxicity test reports; it is not intended to be a treatise on the science or rationale underlying the discipline. The authors have selected for discussion those factors and findings of developmental toxicity tests, the understanding of which they have found to be essential for critical evaluation of the reports of such studies.

Acknowledgement

References

[1] U.S. Environmental Protection Agency, Guidelines for developmental toxicity risk assessment, Fed. Reg., 56 (1991) 63798–63824.

[2] U.S. Food and Drug Administration Guidelines for Reproduction Studies for Safety Evaluation of Drugs for Human Use, Food and Drug Administration, Washington, DC, 1966.

[3] U.S. Environmental Protection Agency, Pesticides Assessment Guidelines, Subdivision F, Hazard Evaluation: Human and Domestic Animals, EPA 450/98-2-025, National Technical Information Service, Springfield, VA, 1982.

[4] (a) U.S. Environmental Protection Agency, Toxic Substances Control Act Test Guidelines. Final rules, Fed. Reg., 50 (1985) 39252–39516. (b) U.S. Environmental Protection Agency, Health Effects Division Standard Evaluation Procedure: Developmental Toxicity Studies (Draft), October 1993. (c) U.S. Environmental Protection Agency, Health Effects Division Standard Evaluation Procedure: Reproductive Toxicity Studies (Draft), October 1993.

[5] Interagency Regulatory Liaison Group, Testing Standards and Guidelines Workgroup, Recommended Guidelines for Teratogenicity Studies in the Rat, Mouse, Hamster or Rabbit, Publication No. PB-82-119 488, National Technical Information Services, Springfield, VA, 1981.

[6] Ministry of Health and Welfare, Canada, The Testing of Chemicals for Carcinogenicity, Mutagenicity and Teratogenicity, Minister of Supply and Services, Canada, Ottawa, 1973.

[7] Organization for Economic Cooperation and Development (OECD), Guideline for testing of chemicals' teratogenicity, 1981.

[8] J.M. Manson and Y.J. Kang, Test methods for assessing female reproductive and developmental toxicology, in: A.W. Hayes (Ed.), Principles and Methods of Toxicology, 2nd edn., Raven, New York. 1989, pp. 311–359.

[9] J.L. Schardein, Chemically Induced Birth Defects, Marcel Dekker, New York, 1985.

[10] J.G. Wilson, Environment and Birth Defects, Academic Press, New York, 1973.

[11] U.S. Environmental Protection Agency, Triethylene glycol monomethyl, monoethyl and monobutyl ethers: Proposed test rule, Fed. Reg., 51 (1986) 17883–17894.

[12] U.S. Environmental Protection Agency, Diethylene glycol butyl ether and diethylene glycol butyl ether acetate: Final test rule, Fed. Reg., 53 (1988) 5932–5953.

[13] U.S. Environmental Protection Agency, Triethylene glycol monomethyl ether: Final test rule, Fed. Reg., 54 (1989) 13472–13477.

[14] U.S. Environmental Protection Agency, Pesticide Assessment Guidelines. Subdivision F. Hazard evaluation: Human and Domestic animals. Addendum 10: Neurotoxicity, Series 81, 82 and 83, EPA 540/09-91-123, Office of Pesticides and Toxic Substances, Washington, DC, 1991.

[15] J.M. DeSesso, Maternal factors in developmental toxicity, Teratogenesis Carcinog. Mutagen., 7 (1987) 225–240.

[16] N. Chernoff, R.J. Kavlock, P.E. Beyer and D. Miller, The potential relationship of maternal toxicity, general stress, and fetal outcome, Teratogenesis Carcinog. Mutagen., 7 (1987) 241–253.

[17] J.L. Schardein, Approaches to defining the relationship of maternal and developmental toxicity, Teratogenesis Carcinog. Mutagen., 7 (1987) 255–271.

[18] K.S. Khera, Maternal toxicity: A possible etiological factor in embryo-fetal deaths and fetal malformations of rodent-rabbit species, Teratology, 31 (1985) 129–153.

[19] K.S. Khera, Maternal toxicity in humans and animals: Effects on fetal development and criteria for detection, Teratogenesis Carcinog. Mutagen., 7 (1987) 287–295.

[20] U.S. Food and Drug Administration, Good laboratory practice regulations for non-clinical laboratory studies, Fed. Reg., 43 (1978) 59985–60025.

[21] European Chemical Industry Ecology and Toxicologic Center, Good laboratory practice, Monograph No. 1. Brussels, Belgium, 1979.

[22] Organization for Economic Co-operation and Development (OECD), Good laboratory practice in the testing of chemicals; Final report of the group of experts on good laboratory practice. Paris, France, 1982.

[23] U.S. Environmental Protection Agency. Toxic substance control; Good laboratory practice standards; Final rule. Fed. Reg., 48 (1983) 53922–53944.

[24] U.S. Environmental Protection Agency. Pesticide programs; Good laboratory practice standards; Final rule, Fed. Reg., 48 (1983) 53946–53969.

[25] National Institutes of Health, Guide for the Care and Use of Laboratory Animals, NIH Publication No. 86-23 U.S. Department of Health and Human Services, Public Health Service, Washington, DC, 1985.

[26] U.S. Department of Agriculture, Animal Welfare Act, Public Law 39-544, Section 6, U.S. Department of Agriculture, Washington, DC, 1970.

[27] U.S. Department of Agriculture, Animal Welfare: Proposed rules, Federal Reg., 55 (1990) 33448–33531.

[28] C.A. Kimmel and C.J.Price, Developmental toxicity studies, in: D.L. Arnold, H.C. Grice and D.R. Krewski (Eds.), Handbook of In Vivo Toxicity Testing, Academic Press, San Diego, CA, 1990, pp. 271–301.

[29] R.E. Staples and V.L. Schnell, Refinements in rapid clearing technique in the KOH-alizarin red S method for fetal bone, Stain Technol., 43 (1968) 61–63.

[30] M. Inouye, Differential staining of cartilage and bone in fetal mouse skeleton by alcian blue and alizarin red S, Cong. Anom., 16 (1976) 171–173.

[31] R.E. Staples, Detection of visceral alterations in mammalian fetuses, Teratology, 9 (1974) 37A–38A.

[32] J.L. Stuckhardt and S.M. Poppe, Fresh visceral examination of rat and rabbit fetuses used in teratogenicity testing, Teratogenesis, Carcinog., Mutagen., 4 (1984) 181–188.

[33] J.G. Wilson, Methods for administering agents and detecting malformations in animals, In: J.G. Wilson and J. Warkany (Eds.), Teratology: Principles and Techniques, University of Chicago Press, Chicago, 1965, pp. 262–277.

[34] A.K. Palmer, Incidence of sporadic malformations, anomalies and variations in random bred laboratory animals, in: D. Neubert, H.J. Merker and T.E. Kwasigroch (Eds.), Methods in Prenatal Toxicology, PSG Publishing, Boston, 1977, pp. 52–71.

[35] C.S. Weil, Selection of the valid number of sampling units and a consideration of their combination in toxicological studies involving reproduction, teratogenesis, and carcinogenesis, Food Cosmet. Toxicol., 8 (1970) 177–182.

[36] D.W. Gaylor, Methods and concepts of biometrics applied to teratology, in: J.G. Wilson and F.C. Fraser (Eds.), Handbook of Teratology, Vol. 1, Plenum Press, New York, 1978, pp. 429–444.

[37] S.C. Gad and C.S. Weil, (eds.), Data analysis applications in toxicology, Statistics and Experimental Design for Toxicologists, [Telford Press] Caldwell, New Jersey, 1986, pp. 147–175.

[38] H. Kalter, Experimental investigation of teratogenic action, Ann. New York Acad. Sci., 123 (1965) 287–294.

[39] V.H. Frankos, FDA perspectives on the use of teratology data for human risk assessment, Fundam. Appl. Toxicol., 5 (1985) 615–625.

[40] K.S. Crump, A new method for determining allowable daily intakes, Fundam. Appl. Toxicol., 4 (1984) 854–871.

[41] C.A. Kimmel and H Zenick, Alternatives to the NOAEL/uncertainty factors (UF) approach for quantitative noncancer risk assessment, Fundam. Appl. Toxicol., 20 (1993) 7–9.